

Abstrak

Churn prediction adalah suatu cara untuk memprediksi pelanggan yang berpotensi untuk *churn*. Data mining khususnya klasifikasi tampaknya dapat menjadi salah satu alternatif solusi dalam membuat model *churn prediction* yang akurat. Namun hasil klasifikasi menjadi tidak akurat disebabkan karena data *churn* bersifat *imbalance*. Kelas data menjadi tidak stabil karena data akan lebih condong ke bagian data yang memiliki komposisi data yang lebih besar.

Salah satu cara untuk menangani permasalahan ini adalah dengan memodifikasi dataset yang digunakan atau yang lebih dikenal dengan metode *sampling*. Metode *sampling* ini meliputi *oversampling*, *undersampling*, dan *combine*.

Analisis yang dilakukan pada tugas akhir ini adalah mengetahui bagaimana pengaruh metode *sampling* yang digunakan terhadap akurasi prediksi data *churn* dengan melakukan penghitungan akurasi model *churn prediction* yang dinyatakan dalam bentuk *lift curve*, *top decile* dan *gini coefficient* serta f-measure untuk penghitungan akurasi prediksi data sebagai data yang *imbalance*.

Hasil yang didapat dari penelitian menunjukkan bahwa metode *sampling* dapat meningkatkan nilai akurasi dalam evaluasi data *churn*. Dengan penerapan metode *sampling*, data *churn* yang memiliki tingkat *imbalance* yang besar dapat diklasifikasi tanpa mengorbankan data minor yang menjadi fokus penelitian. Metode *sampling* yang digunakan juga memiliki hasil evaluasi yang berbeda terhadap dataset sebagai data *churn* dan sebagai data *imbalance*.

Kata kunci: *churn prediction*, *imbalance*, *sampling*, akurasi, evaluasi.